

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁷ : C12N	A2	(11) International Publication Number: WO 00/18881 (43) International Publication Date: 6 April 2000 (06.04.00)
(21) International Application Number: PCT/US99/22776 (22) International Filing Date: 30 September 1999 (30.09.99) (30) Priority Data: 60/102,413 30 September 1998 (30.09.98) US (71) Applicant (for all designated States except US): NEW ENGLAND BIOLABS, INC. [US/US]; 32 Tozer Road, Beverly, MA 01915 (US). (72) Inventors; and (75) Inventors/Applicants (for US only): XU, Ming-Qun [US/CN]; 40 Crescent Road, Hamilton, MA 01982 (US). EVANS, Thomas, C. [US/US]; 1st floor, 68 Albion Street, Somerville, MA 02143 (US). (74) Agent: WILLIAMS, Gregory, D.; New England Biolabs, Inc., 32 Tozer Road, Beverly, MA 01915 (US).		(81) Designated States: CA, JP, US, European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published Without international search report and to be republished upon receipt of that report.
(54) Title: INTEIN MEDIATED PEPTIDE LIGATION		
(57) Abstract		
<p>In accordance with the present invention, there is provided a method for producing a semi-synthetic fusion protein <i>in vitro</i>, comprising the steps of producing a target protein fused to a protein splicing element (an intein) and selectively cleaving the fusion and ligating a synthetic protein or peptide at the C-terminal thioester of the target protein, which overcome many of the disadvantages and problems noted above. Specifically, the present invention has higher yields due to better thiol-induced cleavage with thiol-reagents which have been optimized for the ligation reaction, off-column ligation allows sample concentration and allows the use of less peptide, MESNA is an odorless thiol-reagent for ligation, and Mxe intein is from a bacterial source and often expresses better in bacterial cells. Furthermore, the present invention allows peptides to be directly ligated to the thioester bond formed between an intein and the target protein. The present invention also provides a method for producing a cytotoxic protein, comprising the steps of producing a truncated, inactive form of the protein <i>in vivo</i> which is fused to a protein splicing element, and selectively cleaving the fusion and ligating a synthetic protein or peptide at a C-terminal thioester of the target protein to restore the activity of the native cytotoxic protein. Recombinant vectors for producing such cleavable fusion proteins are also provided.</p>		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

INTEIN MEDIATED PEPTIDE LIGATION

BACKGROUND OF THE INVENTION

5 Genetic engineering is a powerful approach to the manipulation of proteins. However, genetic methodologies are constrained by the use of only naturally coded amino acids. Furthermore, cytotoxic proteins are difficult to obtain by expression and isolation from a living source, since the
10 expression of the toxic protein can result in death of the host.

 To some extent, protocols have been developed to circumvent these problems, for example, total chemical synthesis (Kent, S. B. (1988) *Ann. Rev. Biochem.* 57:957-989),
15 use of misacylated tRNAs (Noren, et al., (1989) *Science* 244:182-188), and semi-synthetic techniques (reviewed in Offord, R. (1987) *Protein Eng.* 1:151-157; Roy. et al. (1994) *Methods in Enzymol.* 231:194-215; Wallace, C. J. (1993) *FASEB* 7:505-515). However, all of these procedures are limited by
20 either the size of the fragment which can be generated or by low reaction yield.

 It would therefore be desirable to develop a high-yield, semi-synthetic technique to allow *in vitro* fusion of a
25 synthetic protein or peptide fragment to an expressed protein without limitation as to the size of the fused fragments.

Likewise, in order to produce cytotoxic proteins, it would be desirable to develop a method of fusing a synthetic fragment, *in vitro*, to an inactive, expressed protein, so as to -
5 restore protein activity post-production from the host.

The modified Sce VMA intein has been used to generate thioester-tagged proteins for use in ligation (Example 19, U.S.S.N. 08/811,492, filed June 16, 1997; Chong, (1996) *J. Biol. Chem.*, 271(36):22159-22168; Chong, (1997) *Gene*, 10 192:271-281; and Muir, et al. (1998) *Proc. Natl. Acad. Sci USA* 95:6705-6710).

Some disadvantages have been low yields due to poor
15 cleavage of the Sce VMA intein with thiol-reagents that are optimum for ligation, the need for large peptide quantities due to on-column reactions, the use of odoriferous reagents, and/or low protein yields due to the use of a large, eukaryotic intein.

SUMMARY OF THE INVENTION

In accordance with the present invention, there is
provided a method for producing a semi-synthetic fusion
25 protein *in vitro*, comprising the steps of producing a target protein fused to a protein splicing element (an intein) and selectively cleaving the fusion and ligating a synthetic

protein or peptide at the C-terminal thioester of the target protein, which overcome many of the disadvantages and problems noted above. Specifically, the present invention has higher yields due to better thiol-induced cleavage with thiol reagents which have been optimized for the ligation reaction. Off-column ligation allows for sample concentration as well as the use of less peptide. In a particularly preferred embodiment, thiol reagents such as 2-mercaptoethanesulfonic acid (MESNA), which is an odorless thiol-reagent, is used for cleavage and ligation along with the Mxe intein, which is from a bacterial source and often expresses better in bacterial cells. Furthermore, the present invention allows peptides to be directly ligated to the thioester bond formed between an intein and the target protein. The present invention also provides a method for producing a cytotoxic protein, comprising the steps of producing a truncated, inactive form of the protein *in vivo* which is fused to a protein splicing element, and selectively cleaving the fusion and ligating a synthetic protein or peptide at a C-terminal thioester of the target protein to restore the activity of the native cytotoxic protein. Recombinant vectors for producing such cleavable fusion proteins are also provided.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 is a flow diagram depicting the chemical reactions which enable intein-mediated peptide ligation. The

thioester generated at the C-terminus of the target protein during IMPACT™ purification was used in a 'native chemical ligation' reaction. This allowed the ligation of a synthetic peptide to a bacterially expressed protein. A typical ligation -
5 reaction involved the expression of the target protein-intein-CBD fusion followed by binding to a chitin resin. A thiol reagent induced cleavage of the intein. The target was eluted from the chitin resin and a synthetic peptide was added. The ligation reaction proceeded overnight.

10 Figure 2 is a gel depicting the results of cleavage and ligation reactions using various thiols. Cleavage and ligation reactions with different thiols visualized on 10-20% Tricine gels. MYB (a fusion protein of maltose binding protein-Sce
15 VMA intein (N454A)-chitin binding domain) and MXB (a fusion protein of maltose binding protein-Mxe GyrA (N198A) intein-chitin binding domain) were incubated overnight at 4°C with various thiols (50 mM) in 150 mM Tris, 100 mM NaCl, pH 8 in the presence of a 30 amino acid peptide with an N-terminal
20 cysteine. The peptide ligates to the C-terminus of MBP. Lanes 1-5 ligation with MYB. Lane 1 no thiol. Lane 2 dithiothreitol. Lane 3 2-mercaptoethanesulfonic acid. Lane 4 3-mercaptopropionic acid. Lane 5 thiophenol. Lanes 6-10
ligation with MXB. Lane 6 no thiol. Lane 7 dithiothreitol. Lane
25 8 2-mercaptoethanesulfonic acid. Lane 9 3-mercaptopropionic acid. Lane 10 thiophenol.

Figure 3 is a gel depicting direct ligation of a peptide to the thioester formed between the Sce VMA intein and maltose binding protein. SDS-PAGE of direct ligation reaction with a 10-20% Tricine gel. Lane 1: a precursor protein (MYBleu) consisting of maltose binding protein-Sce VMA1 intein-chitin binding domain was heated to >95°C for 5 minutes in a buffer of 50 mM Trizma base, pH 8.5 containing 100 mM NaCl, 1% SDS, and mM tris-(2-carboxyethyl)phosphine (TCEP) followed by overnight incubation at room temperature. The precursor (MYBleu) is visible along with the Sce VMA1 intein (Y) and maltose binding protein (M), which are cleavage products. Lane 2: the precursor protein was subjected to the same conditions as described in Lane 1 except that the 30 amino acid peptide (1 mM) was added. The precursor (MYB) and cleavage products (Y and M) are visible along with the ligation product (M+30mer) formed when the 30 amino acid peptide fuses to maltose binding protein.

Figure 4 is a diagram depicting the pTXB1 expression vector of Example I (SEQ ID NO:7 and SEQ ID NO:8).

Figure 5 is the DNA sequence of pTXB1 (SEQ ID NO:5).

Figure 6 is a gel depicting the results of the *HpaI* protein ligation reaction. Protein ligation reactions examined on 10-20% Tricine gels. Lane 1: clarified cells extract after IPTG (0.5 mM) induction of ER2566 cells containing the pTXB2-*HpaI*

plasmid. The fusion protein of *HpaI*₂₂₃-Mxe GyrA-intein-CBD (52 kDa) is visible. Lane 2: cell extract as in Lane 1 after passage over a chitin column, which results in the binding of the fusion protein. Lane 3: *HpaI*₂₂₃ (25.7 kDa) after cleavage
5 from the fusion protein by addition of MESNA. Lane 4: ligation product of *HpaI*₂₂₃ (0.2 mg/mL) with 1 mM of a 31 amino acid peptide (ligation product 29.6 kDa), representing the residues necessary to generate full length *HpaI*, after overnight incubation at 4°C. Lane 5: full length *HpaI* from a recombinant
10 source (29.6 kDa) containing BSA (66 kDa) and two impurities.

Figure 7 is a western blot of various proteins ligated to a biotinylated peptide. Proteins purified with the Mxe GyrA IMPACT™ derivative were ligated to a synthetic peptide which
15 contained an antibody recognition sequence.

DETAILED DESCRIPTION OF THE INVENTION

The ligation methods of the present invention are based
20 on the discovery that a cysteine or peptide fragment containing an N-terminal cysteine may be fused, *in vitro*, to a bacterially expressed protein produced by thiol-induced cleavage of an intein (U.S. Patent No. 5,496,714; Example 19 of U.S.S.N. 08/811,492 filed June 16, 1997; Chong, et al., (1996)
25 *supra* and Chong, et al., (1997) *supra*.

The ligation procedure disclosed herein utilizes a protein splicing element, an intein (Perler, et al., (1994) *Nucleic Acids Res.* 22:1125-1127) to precisely create a thioester at the C-terminal α -carbon of an expressed protein. -

5 This reactive thioester could be present between the target protein and intein or generated by the addition of a thiol reagent. Previously the generation such a thioester was described using an intein (CIVPS) that was modified to undergo thiol inducible cleavage at its N-terminal junction in
10 the presence of thiol reagent dithiothreitol (DTT) (Chong, et al. (1997) *supra*; Comb, et.al. U.S. Patent No. 5,496,714). This C-terminal thioester was previously used in a 'native chemical ligation' type reaction (Dawson, et al., (1994) *Science* 266:776-779) to fuse ³⁵S-cysteine or a peptide
15 fragment containing an N-terminal cysteine to a bacterially expressed protein (Example 19, Comb, et.al. U.S. Patent No. 5,834,247, Chong (1996) *supra* and Chong (1997) *supra*.

20 The ligation method of the instant invention begins with the purification of the thioester-tagged target protein using an intein as described (Chong, et.al. (1997) *supra*). The direct ligation method of the instant invention begins with the isolation of a precursor composed of the target protein-intein-CBD. In one preferred embodiment, the host cell is
25 bacterial. In other embodiments the host cell may be yeast, insect, or mammalian. A cysteine thiol at the N-terminus of a synthetic peptide nucleophilically attacks a thioester present

on the freshly isolated C-terminal α -carbon of the target protein or directly attacks the thioester present between the target protein and intein. This initially generates a thioester between the two reactants which spontaneously rearranges
5 into a native peptide bond (Figure 1).

In order to optimize the ligation efficiency so that greater than 90% of the bacterially expressed target protein can be fused to the synthetic peptide or protein, specific thiol
10 reagents and inteins are screened. In a preferred embodiment, the intein may be any CIVPS, such as *Sce* VMA, *Mxe* GyrA or derivatives of mutants thereof, and the thiol reagent is 2-mercapto-ethanesulfonic acid, thiophenol, DTT, or 3-mercaptopropionic acid (Comb, et al., U.S. Patent No.
15 5,496,714; U.S. Patent No. 5,834,247).

In one particularly preferred embodiment, an intein whose protein splicing activity has been blocked by mutation is utilized. The mutant must, however, retain the ability to
20 undergo the N-S shift, thus allowing thioester formation between itself and an N-terminal protein. This thioester can then be nucleophilically attacked by a thiol reagent or by the N-terminal cysteine of a peptide sequence. For example, by
25 mutating the C-terminal asparagine (asn 198) of an intein from the GyrA gene of *Mycobacterium xenopi* (Telenti, et al., (1997) *J Bacteriol* 179:6378-6382) to an alanine created a

thiol inducible cleavage element. This modified intein cleaved well with thiol reagents that were optimal for the ligation reaction, such as MESNA and thiophenol. Furthermore, optimal thiol reagent and intein combinations can be determined by
5 incubating a precursor protein containing the intein of interest with a wide variety of thiol reagents followed by determination of the extent of cleavage of the precursor protein (Figure 2).

10 The use of such intein and specific thiol reagents leads to optimal yields and high ligation efficiencies; typically greater than 90% of the N-terminal ligation fragment can be modified.

15 The ligation methods of the present invention expand the ability to incorporate non-coded amino acids into large protein sequences by generating a synthetic peptide fragment with fluorescent probes, spin labels, affinity tags, radiolabels, or antigenic determinants and ligating this to an
20 *in vivo* expressed protein isolated using a modified intein.

Furthermore, this procedure allows the isolation of cytotoxic proteins by purifying an inactive truncated precursor from a host source, for example bacteria, and
25 generating an active protein or enzyme after the ligation of a synthetic peptide. For example, restriction endonucleases which have not successfully been cloned by traditional

methods may be produced in accordance with the present invention.

Also, the direct ligation procedure allows the ligation of -
5 a protein or peptide sequence to another protein or peptide
sequence without the use of exogenous thiol reagents. Direct
ligation relies on the nucleophilic attack of the N-terminal
amino acid of one peptide on the thioester formed between a
target protein and an intein (Figure 3).

10 In summary, a fusion protein can be created using the
methods of the present invention that possesses unique
properties which, currently, can not be generated genetically.

15 The Examples presented below are only intended as
specific preferred embodiments of the present invention and
are not intended to limit the scope of the invention. The
present invention encompasses modifications and variations
of the methods taught herein which would be obvious to one of
20 ordinary skill in the art.

The references cited above and below are herein
incorporated by reference.

EXAMPLE I**Creation of vectors pTXB1 and pTXB2 for ligation:**

5 Asparagine 198 of the *Mxe* GyrA intein (Telenti, et al.,
(1997) *J Bacteriol.* 179:6378-6382) was mutated to alanine by
linker insertion into the *XmnI* and *PstI* sites of
pmxeMIPTyrXmnSPdel to create pMXP1. The *XmnI* site was
originally introduced into the unmodified *Mxe* GyrA intein
10 sequence by silent mutagenesis. The *PstI* site was a unique
site in the plasmid. The linker was composed of mxe#3 (5'-
GGTTCGTCAGCCACGCTACTGGCCTCACCGGTTGATAGCTGCA-3')
(SEQ ID NO:1) and mxe#4 (5'-GCTATCAACCGGTGAGGCCAGTAG
CGTGGCTGACGAACC-3') (SEQ ID NO:2).

15 Into pMXP1 another linker composed of mxe#1 (5'-TC
GAATCTAGACATATGGCCATGGGTGGCGGCCGCTCGAGGGCTCTTCC
TGCATCACGGGAGATGCA-3') (SEQ ID NO:3) and mxe#2 (5'-CTAG
TGCATCTCCCGTGATGCAGGAAGAGCCCTCGAGGCGHGCCGCCACCCA
20 TGGCCATATGTCTAGAT-3') (SEQ ID NO:4) was inserted into the
XhoI and *SpeI* sites to introduce a multiple cloning site (*XbaI*-
NdeI-*NcoI*-*NotI*-*XhoI*-*SapI*) before the *Mxe* GyrA intein (pMXP2).

 The 0.6 kilobase *NotI* to *AgeI* fragment of pMXP2 was
25 ligated into the same sites in pTYB1 (IMPACT kit, New England
Biolabs, Beverly, MA) and the *NcoI* to *AgeI* fragment of pMXP2

was cloned into pTYB3 (IMPACT kit, New England Biolabs, Beverly, MA) to create plasmids pTXB1 (see Figure 4 and 5) (SEQ ID NO:5) and pTXB2, respectively. These vectors have a multiple cloning site upstream of the modified *Mxe* GyrA intein-chitin binding domain fusion. This allows the insertion of a target gene of interest inframe with the intein and chitin binding domain (CBD).

Creation of vectors pMYBleu for ligation:

pMYBleu was as described in Chong, et al., (1998), *J. Biol. Chem.* 273:10567-10577. This vector consisted of maltose binding protein upstream of the *Sce* VMA intein-chitin binding domain. A leucine is present at the -1 position instead of the native residue (which is a glycine).

Purification of Thioester-Tagged Proteins:

Protein purification was as described using the *Sce* VMA intein (Chong, et.al., (1997) *Gene* 192:271-281) with slight modification. ER2566 cells (IMPACT T7 instruction manual from New England Biolabs, Beverly, MA) containing the pTXB vector with the appropriate insert were grown to an OD₆₀₀ of 0.5-0.6 at 37°C at which point they were induced with 0.5 mM IPTG overnight at 15°C. Cells were harvested by centrifugation and lysed by sonication (performed on ice). The

three part fusion protein was bound to chitin beads (10 mL bed volume, Figure 6, lanes 1 and 2) equilibrated in Buffer A (50 mM Tris, pH 7.4, and 500 mM NaCl), and washed with 10 column volumes of Buffer A to remove unbound material.

5

Cleavage was initiated using a buffer of 50 mM 2-mercaptoethanesulfonic acid (MESNA), 50 mM Tris, pH 8.0 and 100 mM NaCl. Other thiol reagents were also used at other times, such as thiophenol, dithiothreitol, and/or 3-mercaptopropionic acid. After overnight incubation at from 4-25°C protein was eluted from the column (Figure 6 lane 3). This protein contained a thioester at the C-terminus.

10

Purification of MYB. MYBleu and MXB:

15

Full length precursor proteins consisting of maltose binding protein-*Sce* VMA intein (N454A)-chitin binding domain (MYB) and maltose binding protein-*Mxe* GyrA (N198A) intein-chitin binding domain (MXB) were purified after induction and sonication, as described above, by applying the sonicated sample to a 10 mL column of amylose resin (New England Biolabs, Beverly, MA). Unbound proteins were washed from the column with 10 column volumes of Buffer A (see purification of thioester-tagged proteins). Bound proteins were eluted with a buffer of 50 mM Tris, pH 8, containing 100 mM NaCl and 10 mM maltose. Fractions were collected and protein

20

25

concentrations were determined using the Bio-Rad Protein Assay (Hercules, CA).

Peptide Synthesis:

5

Peptides for subsequent ligation reactions were synthesized on an ABI model 433A peptide synthesizer utilizing *FastMoc*TM chemistry (Fields, et al., (1991) *Pept Res* 4, 95-101) at a 0.085 mmol scale. Preloaded HMP (p-
10 hydroxymethylphenoxymethyl) polystyrene resins (Applied Biosystems, Foster City, CA) functionalized at 0.5 mmol/g was used in conjunction with Fmoc/NMP chemistry utilizing HBTU amino acid activation (Dourtoglou, et al., (1984) *Synthesis* 572-574; Knorr, et al., (1989) *Tetrahedron Lett* 30, 1927-1930). Fmoc amino acids were purchased from Applied
15 Biosystems (Foster City, CA).

Synthesis proceeded with a single coupling during each cycle. Peptide cleavage from the resin and simultaneous
20 removal of side chain protecting groups was facilitated by the addition of cleavage mixture (Perkin Elmer, Norwalk, CT) consisting of 0.75 g phenol, 0.25 mL 1,2-ethanedithiol, 0.5 mL deionized H₂O, and 10 mL TFA. The resin was flushed with nitrogen and gently stirred at room temperature for 3 hours.
25 Following filtration and precipitation into cold (0°C) methyl-t-butyl ether, the precipitate in the ether fraction was

collected by centrifugation. The peptide precipitate was vacuum dried and analyzed by mass spectrometry using a Perceptive Biosystems (Framingham, MA) MALDI-TOF mass spectrometer.

5

Final purification was by HPLC using a Waters HPLC system with a Lambda-Max Model 481 Multiwavelength detector (set at 214 nm), 500 series pumps and automated gradient controller with a Vydac semi-preparative C18 column. Elution of the peptide was with a 60 minute linear gradient of 6-60% acetonitrile (v/v) in an aqueous solution of 0.1% TFA (v/v).

10

Protein Cleavage and Ligation Reactions:

15

Cleavage of MYB and MXB: The precursor protein (1 mg/mL) was incubated overnight at 4°C with or without a thiol reagent (50 mM) in 150 mM Tris, pH 8, containing 100 mM NaCl.

20

Ligation reactions with MYB and MXB: The precursor protein (1 mg/mL) was treated as described for cleavage except that a 30 amino acid peptide (1 mM final concentration, NH₂-CAYKTTQANKHIIIVACEGNPYVPVHFDASV-COOH (SEQ ID NO:6) was also included in the reaction (Figure 2).

25

Ligation reactions after purification of thioester-tagged proteins: Lyophilized peptides (New England Biolabs, Beverly, MA) were added (to 1 mM final concentration) directly to the thioester-tagged protein freshly isolated from the chitin column. The reaction was allowed to proceed overnight at from 4-25°C. In both ligation procedures the condensation of the reactants is visible on a 10-20% Tricine gel (Figure 6). The ligation reaction was tested in conditions of 5-150 mM Tris or HEPES buffers, 50-1000 mM NaCl, 10 mM Maltose, and pH 6-11 and 0-6 M Urea.

Direct Ligation Reactions:

MYBleu (1 mg/mL) was incubated in 6 M Urea or 1% SDS, pH 7.5-8.5, 50-200 mM NaCl, and 1 mM of a 30 amino acid peptide (NH₂CAYKTTQANKHIVVACEGNPYVPVHFDASV-COOH (SEQ ID NO:6)). The MYBleu was incubated for 0-180 minutes at either 4°C or 100°C prior to the addition of the 30 amino acid peptide. Ligation reactions proceeded overnight at either 4°C or 25°C.

EXAMPLE II**Labeling a target protein: Maltose Binding Protein**

5 Maltose binding protein (MBP, 42 kDa) was isolated as described in Example I above using the IMPACT procedure (IMPACT manual from New England Biolabs, Inc., Beverly, MA) in the presence of MESNA.

10 A biotinylated peptide possessing an N-terminal cysteine (CDPEK*DS-COOH (SEQ ID NO:9)), in which the biotin was attached to the ϵ -amino group of the lysine residue) was ligated to the freshly purified target protein as described above. Briefly, 4 μ L of biotinylated peptide (10 mM) were
15 mixed with a 36 μ L aliquot of the freshly purified MBP sample. The mixture was incubated at 4°C overnight.

 Western blots with alkaline phosphatase linked anti-biotin antibody detected the presence of the ligated product
20 but not the unligated target protein (Figure 7). The efficiency of the ligation is typically greater than 90% when MESNA is used for cleavage.

EXAMPLE III**Labeling a target protein: Bst DNA Polymerase I Large Fragment (Bst Pol 1)**

5

Bst DNA Polymerase I large fragment (67 kDa) was isolated as described in Example I above using the IMPACT procedure (IMPACT manual from New England Biolabs, Inc., Beverly, MA) in the presence of MESNA.

10

A biotinylated peptide possessing an N-terminal cysteine (CDPEK*DS-COOH (SEQ ID NO:9)), in which the biotin was attached to the ϵ -amino group of the lysine residue) was ligated to the freshly purified target protein as described.

15

Briefly, 4 μ L of biotinylated peptide (10 mM) were mixed with a 36 μ L aliquot of the freshly purified Bst Pol 1 sample. The mixture was incubated at 4°C overnight.

20

Western blots with alkaline phosphatase linked anti-biotin antibody detected the presence of the ligated product but not the unligated target protein (Figure 7). The efficiency of the ligation is typically greater than 90% when MESNA is used for cleavage.

EXAMPLE IV

Labeling a target protein: Paramyosin

5 Paramyosin (29 kDa) was isolated as described in Example I above using the IMPACT procedure (IMPACT manual from New England Biolabs, Inc., Beverly, MA) in the presence of MESNA.

10 A biotinylated peptide possessing an N-terminal cysteine (CDPEK*DS-COOH (SEQ ID NO:9)), in which the biotin was attached to the ϵ -amino group of the lysine residue) was ligated to the freshly purified target protein as described. Briefly, 4 μ L of biotinylated peptide (10 mM) were mixed with
15 a 36 μ L aliquot of the freshly purified paramyosin sample. The mixture was incubated at 4°C overnight.

 Western blots with alkaline phosphatase linked anti-biotin antibody detected the presence of the ligated product
20 but not the unligated target protein (Figure 7). The efficiency of the ligation is typically greater than 90% when MESNA is used for cleavage.

EXAMPLE V**Labeling a target protein: *E. coli* Thioredoxin**

5 *E. coli* thioredoxin (12 kDa) was isolated as described in Example I above using the IMPACT procedure (IMPACT manual from New England Biolabs, Inc., Beverly, MA) in the presence of MESNA.

10 A biotinylated peptide possessing an N-terminal cysteine (CDPEK*DS-COOH (SEQ ID NO:9)), in which the biotin was attached to the ϵ -amino group of the lysine residue) was ligated to the freshly purified target protein as described. Briefly, 4 μ L of biotinylated peptide (10 mM) were mixed with
15 a 36 μ L aliquot of the freshly purified thioredoxin sample. The mixture was incubated at 4°C overnight.

20 Western blots with alkaline phosphatase linked anti-biotin antibody detected the presence of the ligated product but not the unligated target protein (Figure 7). The efficiency of the ligation is typically greater than 90% when MESNA is used for cleavage.

EXAMPLE VI**Isolation of a cytotoxic protein:**

5 The ligation procedure of Example I was applied to the isolation of a potentially cytotoxic protein. An endonuclease from *Haemophilus parainfluenzae* (*HpaI*; Ito, et al., (1992) *Nucleic Acids Res* 20:705-709) was generated by ligating an inactive truncated form of the enzyme expressed in *E. coli*
10 (ER2566 cells, New England Biolabs, Inc., Beverly, MA) with the missing amino acids that were synthesized chemically.

 The first 223 amino acids of *HpaI* (full length *HpaI* is 254 amino acids) were fused in frame with the modified *Mxe*
15 GyrA intein and the CBD. The 223 amino acid *HpaI* fragment was isolated as described for purification of thioester tagged proteins. The truncated *HpaI* displayed no detectable enzymatic activity.

20 A synthetic peptide representing the 31 amino acids needed to complete *HpaI* was ligated onto the 223 amino acid truncated form of *HpaI* by the method of Example I.

Enzymatic Assay for *Hpa*I:

5 The activity of the fused *Hpa*I was determined by its
ability to digest Lambda DNA (New England Biolabs, Beverly,
MA). Serial dilutions of ligated or truncated *Hpa*I, with the
appropriate peptide added to 1 mM, were incubated with 1 μ g
of Lambda DNA for 1 hour at 37°C in a buffer of 20 mM Tris-
acetate, pH 7.9, 10 mM magnesium acetate, 50 mM potassium
acetate, 1 mM dithiothreitol, and 170 μ g/mL BSA (total
10 volume 30 μ L). Digestion reactions were visualized on 1%
agarose gels permeated with ethidium bromide. One unit of
Hpa I was defined as the amount of enzyme necessary to
digest 1 μ g of Lambda DNA in one hour at 37°C.

15 The newly ligated *Hpa*I had a specific activity of 0.5-
1.5x10⁶ units/mg which correlated well with the expected
value of 1-2x10⁶ units/mg for the full length enzyme.

WHAT IS CLAIMED IS:

1. A method for fusing an expressed protein with a peptide,
said method comprising the steps of:
 - 5 (a) generating at least one C-terminal thioester-tagged target protein;
 - (b) generating at least one target peptide having a specified N-terminal; and
 - (c) ligating said target peptide to said target protein.
- 10 2. The method of claim 1, wherein said target protein is generated from a first plasmid comprising an intein having N-terminal cleavage activity.
- 15 3. The method of claim 2, wherein said intein comprises an intein having a cysteine residue at the N-terminal of the intein.
- 20 4. The method of claim 3, wherein said target protein is generated by thiol reagent-induced cleavage of said intein.
- 25 5. The method of claim 4, wherein said thiol reagent is selected from the group consisting of MESNA, thiophenol, DTT, β -mercaptoethanol or derivatives thereof.
6. A fusion protein produced by the method of any one of claims 1-5.

7. A cyclic protein produced by the method of claim 1.
- 5 8. A modified intein comprising a mutant Mxe GyrA intein -
capable of thiol reagent-induced cleavage to produce a
thioester at the C-terminal of an adjacent target
protein.
- 10 9. A method of generating a reactive thioester comprising
contacting a thiol reagent selected from the group
consisting essentially of MESNA, thiophenol, DTT, β -
mercaptoethanol or derivatives thereof with a precursor
comprising a target protein and intein.
- 15 10. A method for screening thiol reagents which cleave a
target intein comprising the steps of:
 - (a) isolating a precursor comprising a protein and a
modified intein;
 - 20 (b) contacting a thiol reagent with the precursor of
step (a);
 - (c) determining whether a splicing or cleaving event
occurs.
- 25 11. The method of claim 10, comprising the further step of
determining whether the spliced or cleaved product of
step (c) can ligate to a target peptide having an N-
terminal cytokine.

Figure 1
Intein Mediated Protein Ligation

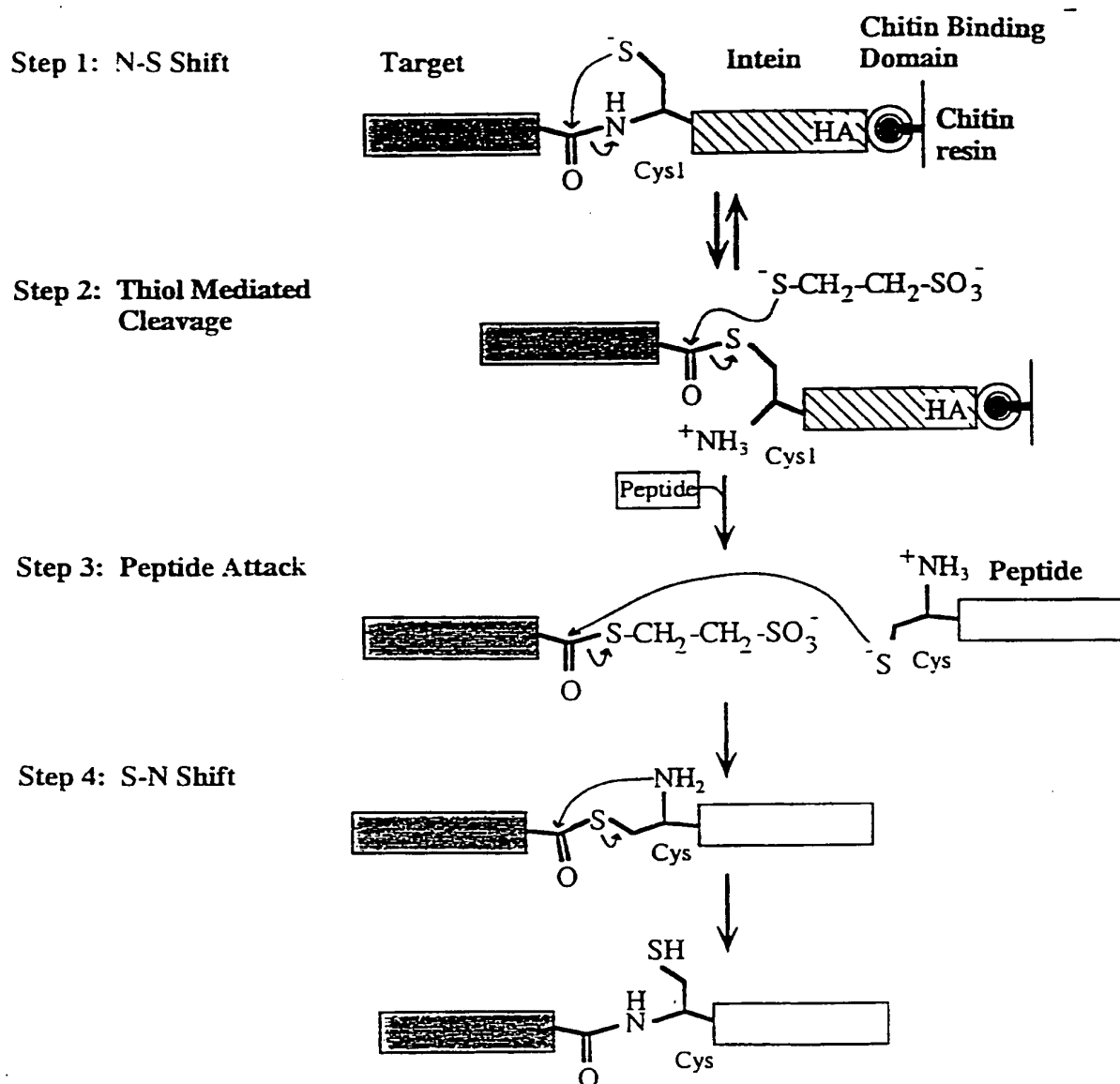


Figure 2

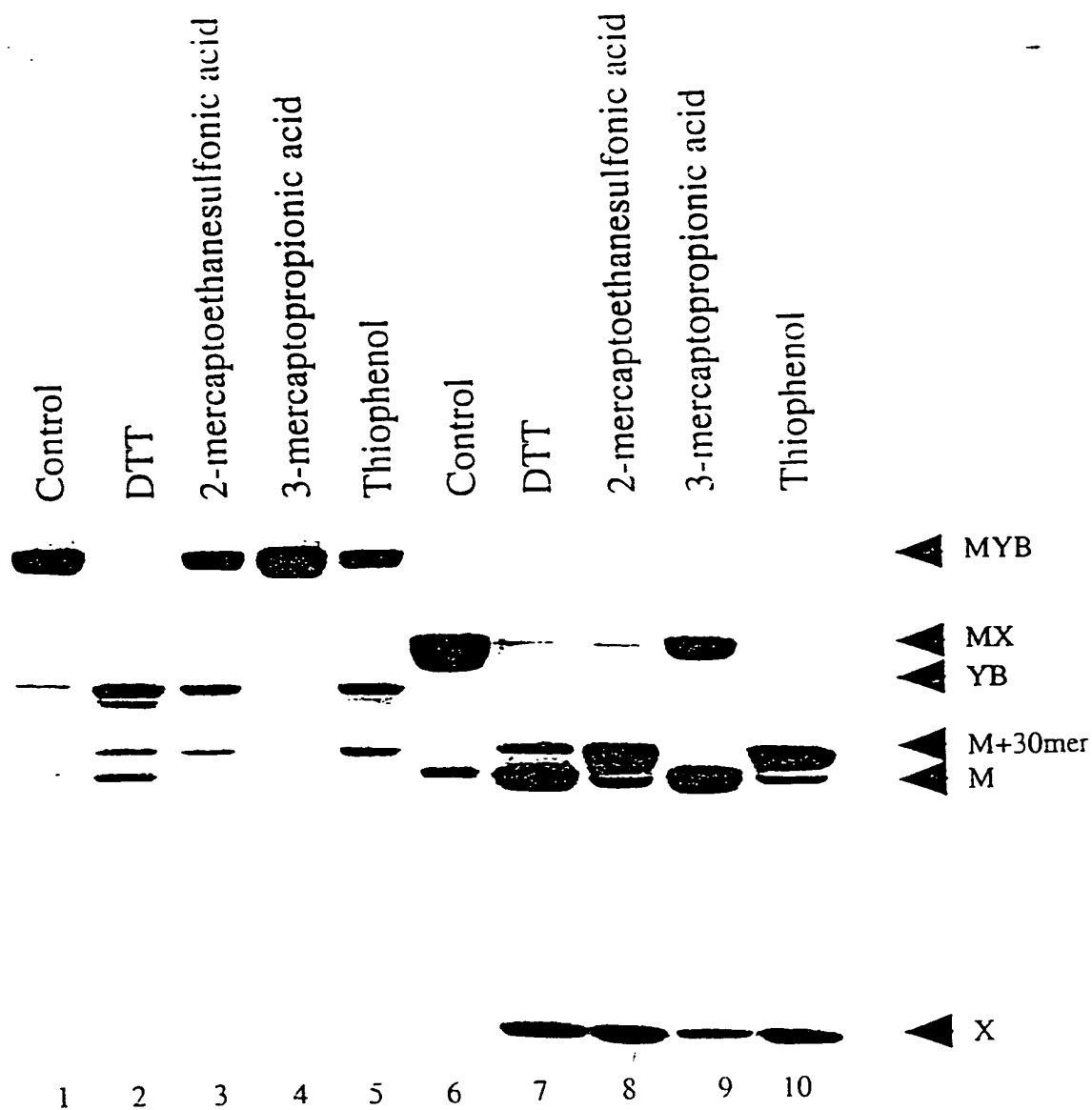
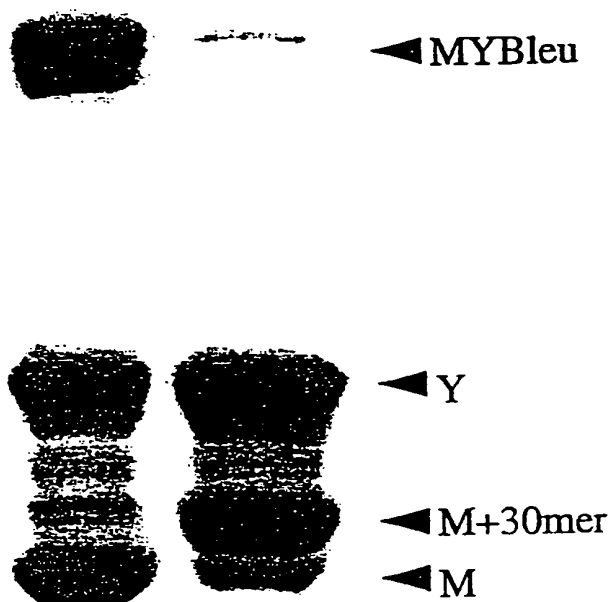
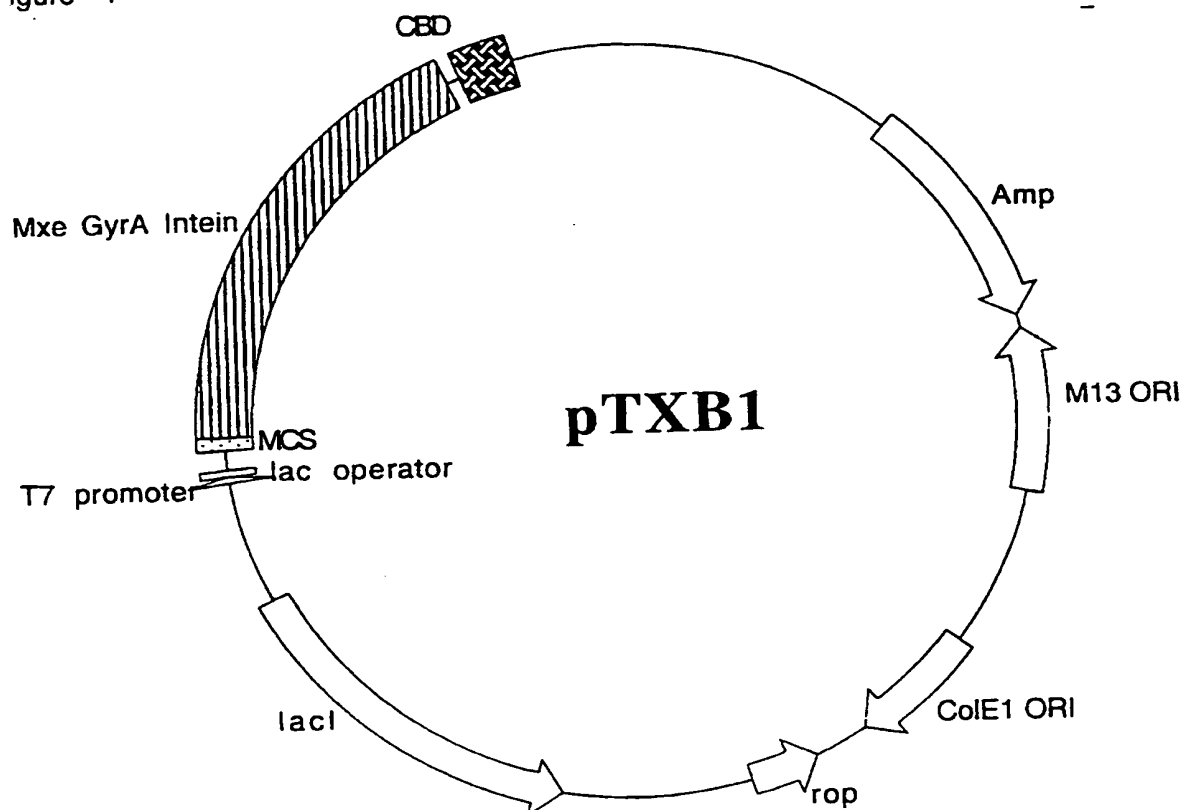


Figure 3. Direct Ligation Reaction



pTXB1 Expression Vector

Figure 4



pTXB1
Met Ala Met Gly Gly Gly Arg Leu Glu Gly Ser Ser Cys1...Intein
CAT ATG GCC ATG GGT GGC GGC CGC CTC GAG GGC TCT TCC TGC
 NdeI NcoI NotI XhoI SapI

Figure 5

1 AACTACGTCA GGTGGCACTT TTCGGGGAAA TGTGCGCGGA ACCCCTATTT
51 GTTTATTTTT CTAAATACAT TCAAATATGT ATCCGCTCAT GAGACAATAA
101 CCCTGATAAA TGCTTCAATA ATATTGAAAA AGGAAGAGTA TGAGTATTCA
151 ACATTTCCGT GTCGCCCTTA TTCCCTTTTT TCGGGCATTT TGCCTTCCTG
201 TTTTGTCTCA CCCAGAAACG CTGGTGAAAG TAAAAGATGC TGAAGATCAG
251 TTGGGTGCAC GAGTGGGTTA CATCGAACTG GATCTCAACA GCGGTAAGAT
301 CCTTGAGAGT TTTGCCCCG AAGAACGTTT TCCAATGATG AGCACTTTTA
351 AAGTTCTGCT ATGTGGCGCG GTATTATCCC GTGTTGACGC CGGGCAAGAG
401 CAACTCGGTC GCCGCATACA CTATTCTCAG AATGACTTGG TTGAGTACTC
451 ACCAGTCACA GAAAAGCATC TTACGGATGG CATGACAGTA AGAGAATTAT
501 GCAGTGCTGC CATAACCATG AGTGATAACA CTGCGGCCAA CTTACTTCTG
551 ACAACGATCG GAGGACCGAA GGAGCTAACC GCTTTTTTGC ACAACATGGG
601 GGATCATGTA ACTCGCCTTG ATCGTTGGGA ACCGGAGCTG AATGAAGCCA
651 TACCAAACGA CGAGCGTGAC ACCACGATGC CTGTAGCAAT GGCAACAACG
701 TTGCGCAAAC TATTAAGTGG CGAACTACTT ACTCTAGCTT CCCGGCAACA
751 ATTAATAGAC TGGATGGAGG CGGATAAAGT TGCAGGACCA CTTCTGCGCT
801 CGGCCCTTCC GGCTGGCTGG TTTATTGCTG ATAAATCTGG AGCCGGTGAG
851 CGTGGGTCTC GCGGTATCAT TGCAGCACTG GGGCCAGATG GTAAGCCCTC

901 CCGTATCGTA GTTATCTACA CGACGGGGAG TCAGGCAACT ATGGATGAAC
951 GAAATAGACA GATCGCTGAG ATAGGTGCCT CACTGATTAA GCATTGGTAA
1001 CTGTCAGACC AAGTTTACTC ATATATACTT TAGATTGATT TACCCCGGTT
1051 GATAATCAGA AAAGCCCCAA AAACAGGAAG ATTGTATAAG CAAATATTTA
1101 AATTGTAAAC GTTAATATTT TGTAAAATT CGCGTTAAAT TTTTGTAA
1151 TCAGCTCATT TTTTAACCAA TAGGCCGAAA TCGGCAAAAT CCCTTATAAA
1201 TCAAAAGAAT AGCCCGAGAT AGGGTTGAGT GTTGTTCAG TTTGGAACAA
1251 GAGTCCACTA TTAAAGAACG TGGACTCCAA CGTCAAAGGG CGAAAAACCG
1301 TCTATCAGGG CGATGGCCCA CTACGTGAAC CATCACCCAA ATCAAGTTTT
1351 TTGGGGTCGA GGTGCCGTAA AGCACTAAAT CGGAACCTA AAGGGAGCCC
1401 CCGATTTAGA GCTTGACGGG GAAAGCCGGC GAACGTGGCG AGAAAGGAAG
1451 GGAAGAAAGC GAAAGGAGCG GCGCTAGGG CGCTGGCAAG TGTAGCGGTC
1501 ACGCTGCGCG TAACCACCAC ACCCGCCGCG CTTAATGCGC CGCTACAGGG
1551 CGCGTAAAAG GATCTAGGTG AAGATCCTTT TTGATAATCT CATGACCAAA
1601 ATCCCTTAAC GTGAGTTTTC GTTCCACTGA GCGTCAGACC CCGTAGAAAA
1651 GATCAAAGGA TCTTCTTGAG ATCCTTTTTT TCTGCGCGTA ATCTGCTGCT
1701 TGCAAACAAA AAAACCACCG CTACCAGCGG TGGTTTGT TTGCCGATCAA
1751 GAGCTACCAA CTCTTTTCC GAAGGTA ACT GGCTTCAGCA GAGCGCAGAT
1801 ACCAAATACT GTCCTTCTAG TGTAGCCGTA GTTAGCCAC CACTTCAAGA
1851 ACTCTGTAGC ACCGCCTACA TACCTCGCTC TGCTAATCCT GTTACCAGTG
1901 GCTGCTGCCA GTGGCGATAA GTCGTGTCTT ACCGGGTTGG ACTCAAGACG
1951 ATAGTTACCG GATAAGGCGC AGCGGTCGGG CTGAACGGGG GGTTCGTGCA
2001 CACAGCCCAG CTTGGAGCGA ACGACCTACA CCGAACTGAG ATACCTACAG
2051 CGTGAGCTAT GAGAAAGCGC CACGCTTCCC GAAGGGAGAA AGGCGGACAG
2101 GTATCCGGTA AGCGGCAGGG TCGGAACAGG AGAGCGCACG AGGGAGCTTC
2151 CAGGGGAAA CGCCTGGTAT CTTTATAGTC CTGTGGGTT TCGCCACCTC
2201 TGA CTGAGC GTCGATTTTT GTGATGCTCG TCAGGGGGG GGAGCCTATG
2251 GAAAAACGCC AGCAACGCGG CCTTTTACG GTTCTGGCC TTTTGCTGGC
2301 CTTTGTCTCA CATGTTCTTT CCTGCGTTAT CCCCTGATTC TGTGGATAAC

2351 CGTATTACCG CCTTTGAGTG AGCTGATACC GCTCGCCGCA GCCGAACGAC
2401 CGAGCGCAGC GAGTCAGTGA GCGAGGAAGC TATGGTGCAC TCTCAGTACA
2451 ATCTGCTCTG ATGCCGCATA GTTAAGCCAG TATACACTCC GCTATCGCTA
2501 CGTGACTGGG TCATGGCTGC GCCCCGACAC CCGCCAACAC CCGCTGACGC
2551 GCCCTGACGG GCTTGTCTGC TCCCGGCATC CGCTTACAGA CAAGCTGTGA
2601 CCGTCTCCGG GAGCTGCATG TGTCAGAGGT TTTCACCGTC ATCACCGAAA
2651 CGCGCGAGGC AGCTGCGGTA AAGCTCATCA GCGTGGTCGT GCAGCGATTG
2701 ACAGATGTCT GCCTGTTTCT CCCTGCTCCAG CTCGTTGAGT TTCTCCAGAA
2751 GCGTTAATGT CTGGCTTCTG ATAAAGCGGG CCATGTTAAG GCGGTTTTTT
2801 TCCTGTTTGG TCACTTGATG CCTCCGTGTA AGGGGGAATT TCTGTTTATG
2851 GGGGTAATGA TACCGATGAA ACGAGAGAGG ATGCTCACGA TACGGGTTAC
2901 TGATGATGAA CATGCCCGGT TACTGGAACG TTGTGAGGGT AAACAACCTGG
2951 CGGTATGGAT GCGGCGGGAC CAGAGAAAAA TCACTCAGGG TCAATGCCag
3001 ccgaACGCCA GCAAGACGTA GCCCAGCGCG TCGGCCGCCA TGCCGGCGAT
3051 AATGGCCTGC TTCTCGCCGA AACGTTTGGT GGCGGGACCA GTGACGAAGG
3101 CTTGAGCGAG GCGGTGCAAG ATTCCGAATA CCGCAAGCGA CAGGCCGATC
3151 ATCGTCGCGC TCCAGCGAAA GCGGTCCTCG CCGAAAATGA CCCAGAGCGC
3201 TGCCGGCACC TGTCTTACGA GTTGCATGAT AAAGAAGACA GTCATAAGTG
3251 CGGCGACGAT AGTCATGCCC CGCGCCCACC GGAAGGAGCT GACTGGGTTG
3301 AAGGCTCTCA AGGGCATCGG TCGAGATCCC GGTGCCTAAT GAGTGAGCTA
3351 ACTTACATTA ATTGCGTTGC GCTCACTGCC CGCTTTCCAG TCGGGAAACC
3401 TGTCGTGCCA GCTGCATTAA TGAATCGGCC AACGCGCGG GAGAGGCGGT
3451 TTGCGTATTG GCGGCCAGGG TGGTTTTTCT TTTCACCACT GAGACGGGCA
3501 ACAGCTGATT GCCCTTCACC GCCTGGCCCT GAGAGAGTTG CAGCAAGCGG
3551 TCCACGCTGG TTTGCCCCAG CAGGCGAAAA TCCTGTTTGA TGGTGGTTAA
3601 CGGCGGGATA TAACATGAGC TGTCTTCGGT ATCGTCGTAT CCCACTACCG
3651 AGATATCCGC ACCAACGCGC AGCCCGGACT CGGTAATGGC GCGCATTGCG
3701 CCCAGCGCCA TCTGATCGTT GGCAACCAGC ATCGCAGTGG GAACGATGCC
3751 CTCATTGAGC ATTTGCATGG TTTGTTGAAA ACCGGACATG GCACTCCAGT

3801 CGCCTTCCCG TTCCGCTATC GGCTGAATTT GATTGCGAGT GAGATATTTA
3851 TGCCAGCCAG CCAGACGCAG ACGCGCCGAG ACAGAACTTA ATGGGCCCCG
3901 TAACAGCGCG ATTTGCTGGT GACCCAATGC GACCAGATGC TCCACGCCCA
3951 GTCGCGTACC GTCTTCATGG GAGAAAATAA TACTGTTGAT GGGTGTCTGG
4001 TCAGAGACAT CAAGAAATAA CGCCGGAACA TTAGTGCAGG CAGCTTCCAC
4051 AGCAATGGCA TCCTGGTCAT CCAGCGGATA GTTAATGATC AGCCCACTGA
4101 CGCGTTGCGC GAGAAGATTG TGCACCGCCG CTTTACAGGC TTCGACGCCG
4151 CTTCTGTTCTA CCATCGACAC CACCACGCTG GCACCCAGTT GATCGGCGCG
4201 AGATTTAATC GCCGCGACAA TTTGCGACGG CGCGTGCAGG GCCAGACTGG
4251 AGGTGGCAAC GCCAATCAGC AACGACTGTT TGCCCCCCAG TTGTTGTGCC
4301 ACGCGGTTGG GAATGTAATT CAGCTCCGCC ATCGCCGCTT CCACTTTTTC
4351 CCGCGTTTTT GCAGAAACGT GGCTGGCCTG GTTCACCACG CGGGAAACGG
4401 TCTGATAAGA GACACCGGCA TACTCTGCGA CATCGTATAA CGTTACTGGT
4451 TTCACATTCA CCACCCTGAA TTGACTCTCT TCCGGGCGCT ATCATGCCAT
4501 ACCGCGAAAG GTTTTGCGCC ATTCGATGGT GTCCCGGATC TCGACGCTCT
4551 CCCTTATGCG ACTCCTGCAT TAGGAAGCAG CCCAGTAGTA GGTGAGGCC
4601 GTTGAGCACC GCCGCCGCAA GGAATGGTGC ATGCCGCCCT TTCGTCTTCA
4651 AGAATTAATT CCAATTCCA GGCATCAAAT AAAACGAAAG GCTCAGTCGA
4701 AAGACTGGGC CTTTCGTTTT ATCTGTTGTT TGTCGGTGAA CGCTCTCCTG
4751 AGTAGGACAA ATCCGCCGGG AGCGGATTG AACGTTGCGA AGCAACGGCC
4801 CGGAGGGTGG CGGGCAGGAC GCCCCCATA AACTGCCAGG AATTAATTCC
4851 AGGCATCAAA TAAAACGAAA GGCTCAGTCG AAAGACTGGG CCTTTCGTTT
4901 TATCTGTTGT TTGTCGGTGA ACGCTCTCCT GAGTAGGACA AATCCGCCGG
4951 GAGCGGATTT GAACGTTGCG AAGCAACGGC CCGGAGGGTG GCGGGCAGGA
5001 CGCCCCCAT AACTGCCAG GAATTAATTC CAGGCATCAA ATAAAACGAA
5051 AGGCTCAGTC GAAAGACTGG GCCTTTCGTT TTATCTGTTG TTTGTCGGTG
5101 AACGCTCTCC TGAGTAGGAC AAATCCGCCG GGAGCGGATT TGAACGTTGC
5151 GAAGCAACGG CCCGGAGGGT GGCGGGCAGG ACGCCCCCA TAACTGCCA
5201 GGAATTAATT CCAGGCATCA AATAAACGA AAGGCTCAGT CGAAAGACTG

5251 GGCTTTTCGT TTTATCTGTT GTTTGTCGGT GAACGCTCTC CTGAGTAGGA
5301 CAAATCCGCC GGGAGCGGAT TTGAACGTTG CGAAGCAACG GCGCGGAGGG
5351 TGGCGGGCAG GACGCCCCGCC ATAAACTGCC AGGAATTGGG GATCGGAATT
5401 AATTCCCGGT TTAAACCGG GATCTCGATC CCGCGAAATT AATACGACTC
5451 ACTATAGGGG AATTGTGAGC GGATAACAAT TCCCCTCTAG AAATAATTTT
5501 GTTTAACTTT AAGAAGGAGA TATAcatatg gctagctcgc gagtcgacgg
5551 cggccgcctc gagggctctt ccTGCATCAC GGGAGATGCA CTAGTTGCCC
5601 TACCCGAGGG CGAGTCGGTA CGCATCGCCG ACATCGTGCC GGGTGCGCGG
5651 CCCAACAGTG ACAACGCCAT CGACCTGAAA GTCCTTGACC GGCATGGCAA
5701 TCCCGTGCTC GCCGACCGGC TGTTCCTCTC CGGCGAGCAT CCGGTGTACA
5751 CCGTGCGTAC GGTCGAAGGT CTGCGTGTGA CGGGCACCGC GAACCAACCG
5801 TTGTGTGTT TGGTCGACGT CGCCGGGGTG CCGACCCTGC TGTGGAAGCT
5851 GATCGACGAA ATCAAGCCGG GCGATTACGC GGTGATTCAA CGCAGCGCAT
5901 TCAGCGTCGA CTGTGCAGGT TTTGCCCGCG GAAAACCCGA ATTTGCGCCC
5951 ACAACCTACA CAGTCGGCGT CCCTGGACTG GTGCGTTTCT TGGAAGCACA
6001 CCACCGAGAC CCGGACGCCC AAGCTATCGC CGACGAGCTG ACCGACGGGC
6051 GGTTCTACTA CGCGAAAGTC GCCAGTGTCA CCGACGCCGG CGTGCAGCCG
6101 GTGTATAGCC TTCGTGTCGA CACGGCAGAC CACGCGTTTA TCACGAACGG
6151 GTTCGTCAGC CACGCTACTG GCCTCACC GG TCTGAACTCA GGCCTCACGA
6201 CAAATCCTGG TGTATCCGCT TGGCAGGTCA ACACAGCTTA TACTGCGGGA
6251 CAATTGGTCA CATATAACGG CAAGACGTAT AAATGTTTGC AGCCCCACAC
6301 CTCCTTGCCA GGATGGGAAC CATCCAACGT TCCTGCCTTG TGGCAGCTTC
6351 AATGactgca ggaaggGGAT CCGGCTGCTA ACAAAGCCCG AAAGGAAGCT
6401 GAGTTGGCTG CTGCCACCGC TGAGCAATAA CTAGCATAAC CCCTTGGGGC
6451 CTCTAAACGG GTCTTGAGGG GTTTTTTGCT GAAAGGAGGA ACTATATCCG
6501 GAT

Figure 6
Hpa I ligation

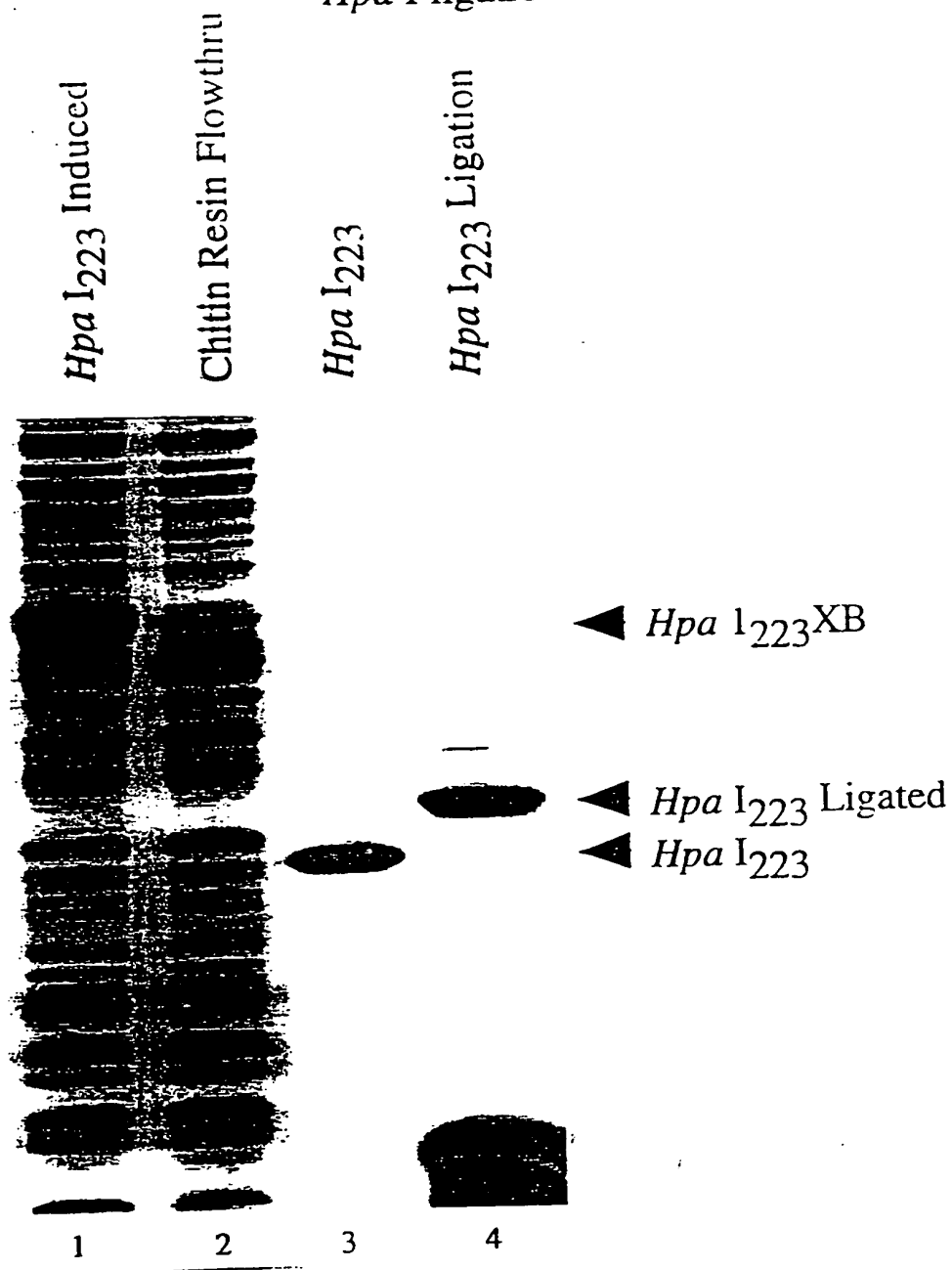


Figure 7

Western blots of Proteins Ligated to a Biotinylated Peptide

MBP+Peptide
MBP Control
Bst Pol I+Peptide
Bst Pol I Control
Paramyosin+Peptide
Paramyosin Control
Thioredoxin+Peptide
Thioredoxin Control

